

Weekly Report

2018.0903-2018.0909

1.This Week

Power Grid Deep Learning Paper

1. We spent 6 days to generate the entire 2,000 node dataset after we finished processing with the first batch. We generate 100,000 samples in total, including normal samples.
2. We generate a small data set of faults happened on different parts of the bus and plan to use the model to classify these sample to see if these kind of data can be classified directly.
3. I talked to Huang this week and he suggests that: it is a good idea to realize automatic parameter recommendation. The recommended parameter set don't have to be the one that can achieve the best performance, they can accept with a generally good set.
4. I read the TPFlow paper and plan to apply this method to our dataset.
5. We originally planed to go to Beijing next week. But Huang said that the program was delayed to October. So we will go in October (if the time conflicts with the VIS, then Chen Zexian need to go alone.)

WaveLines Revision:

- 1.Finish the user study part in the paper and simplify the expression of expert interview. Also, the left minor problems are also revised. One reviewer says we always claim something that we didn't do in our paper and I carefully checked our paper and revise the expressions that might cause such an understanding.
2. We change the color map in the topology view according to the advice from Zhang Wei. We re-made the pictures used in the paper. And We plan to re-made the video the next Monday.

SQC Paper

- 1.Talk to Zongzhuang this week. He is learning d3 right now. I asked him to follow the coding training from Prof. Wu. When he feels he is more familiar with coding, then We discuss about the next step.

Others

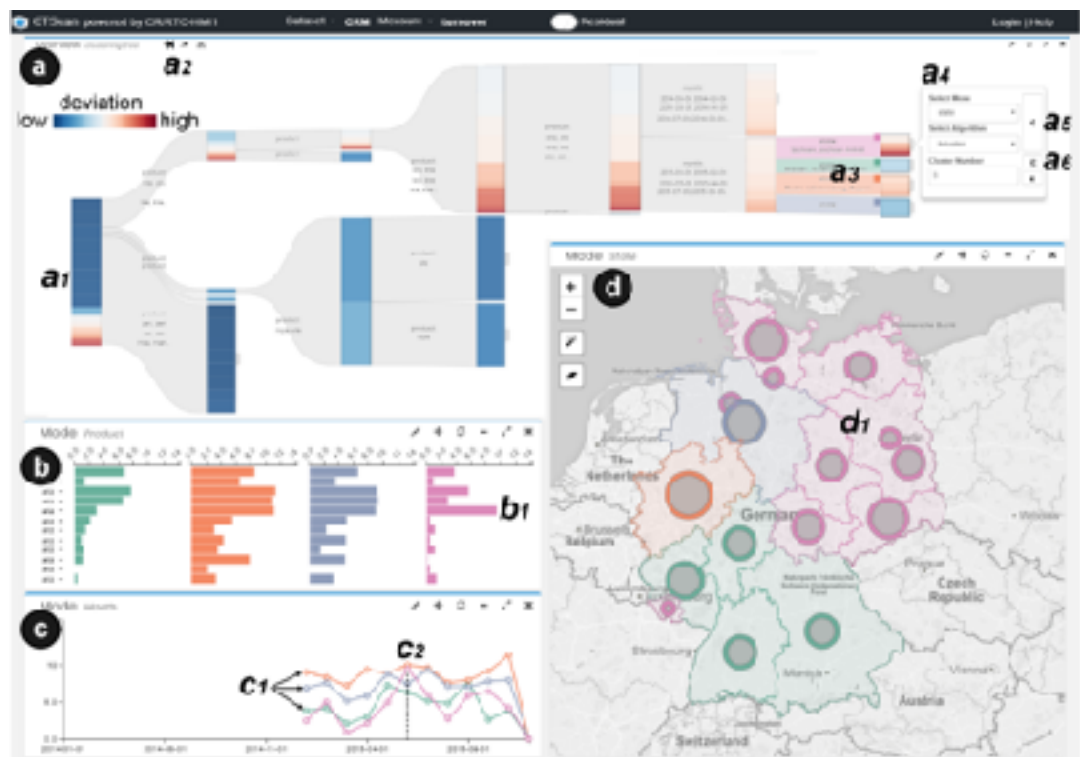
- 1.Place the paper writing tricks (套路) on google docs.
- 2.Read the paper that I'm going to present on the group meeting on 9.17.

Working Hour: (except nap and eat time)
8 - 9 hours / week day
2 - 8 hours / weekend day
Total Working Hour this week: 53 hours.

Paper Reading

1.TPFlow: Progressive Partition and Multidimensional Pattern Extraction for Large-Scale Spatio-Temporal Data Analysis (VIS 2018)

This paper first attracts me by its idea: automatic pattern extraction for spatio-temporal datasets. This is a quite difficult problem that we used want to solve when we are doing the Wavelines paper (But we didn't find a possible solution). This paper used a rank-one tensor decomposition algorithm to realize both dataset partition and multidimensional pattern extraction. Particularly, this algorithm considers how faithfully the extracted patterns can visually represent the original data (which is a very important problem but is always omitted by us). The algorithm it self is simple but clever and the visual design is also simple but exquisite.



2.Trajectory-driven Influential Billboard Placement (KDD 2018)

This paper tries to solve the problem of billboard placement. It first partition the billboard set U into a set of small clusters. Then use a greedy method (enumeration) to compute the the locally influential billboards for each cluster,

which reduces the computation cost greatly. Then billboards with low marginal influence are pruned and the final places are recommended. Last year, Weng Di did a paper about billboard placement with visual approaches. Actually, the visual methods can be further enhanced by such automatic recommendation methods.

3.Exploring the Urban Region-of-Interest through the Analysis of Online Map Search Queries (KDD 2018)

This paper divides the urban area into small region grids. Then it calculates the PageRank values of the areas as visiting popularity based on the transition information extracted from map queries. A density-based clustering method is used to merge neighboring region grids into integrated ROIs. A topic model is then used to identify the latent travel patterns of ROI visitors. Methods used in this paper and the framework are all general methods in data visualization. But this paper performs better than usual data visualization papers studying ROI analysis is that it not only uses the topic model to explain which which ROI attracts people to visit but also why.

4.NetWalk: A Flexible Deep Embedding Approach for Anomaly Detection in Dynamic Networks (KDD 2018)

This is a very first paper that I have learned about anomaly detection in dynamic networks. This paper defines anomaly in a graph as vertices and edges whose “behaviors” deviate from underlying majority of the network. This paper encodes the vertices of the dynamic network to vector representations by clique embedding (this minimizes the distance of vertex representations of each walk of the dynamic network). And then it detects the anomaly by applying a cluster method. I feel that anomaly detection in graph data is something we can do in the future. (Especially dynamic networks)

5.Multi-label Learning with Highly Incomplete Data via Collaborative Embedding (KDD 2018)

This paper uses collaborative embedding to support the coupling process of reconstructing missing features and weak label assignments to solve the incomplete data problem in multi-label learning problems. This method can be used to enhance the performance the of our fault detection model in theory.

2.Progress

Work	Deadline	Progress
Power grid paper with Deeping learning	-	1.Train the model on the 2,000 node dataset.
SQC Paper	-	1.Talk to Zongzhuang this week.
WaveLine revision	ASAP	1.Finished paper revision. 2.Record the video the next week.

